

Business Intelligence et KPI

2. Exemples de cas pratiques

Salim Lardjane

Université de Bretagne Sud

Santé

- On va présenter deux exemples d'application au domaine de la Santé : le traitement de cancers de la peau et le traitement de la pré-éclampsie (hypertension) de la grossesse.

Santé

- **Entreprise** : on souhaite analyser le processus de traitement du cancer de la peau.
- **Objectifs** : ceux-ci se réfèrent au traitement des patients et à la performance de l'Hôpital.
- En ce qui concerne le traitement des patients, les KPI suivants ont été identifiés :

Santé

- Durée de Survie des patients
- Suivi de check-up préventifs par les patients
- En ce qui concerne la performance de l'institution, on considère les KPI suivants :
- Conformité de l'institution aux normes internationales
- Bonne organisation des processus internes

Santé

- **Activité data** : les sources de données incluent l'Hôpital, un registre national du cancer et des données de l'Assurance Maladie. Les données sont diverses en termes de granularité temporelle, de qualité et d'exhaustivité, ainsi qu'en termes de structure (données structurées et semi-structurées).
- On a donc besoin d'un **modèle d'information flexible** qui permette l'intégration de toutes les données en vue des objectifs posés.

Santé

- **Activité de compréhension de l'entreprise et des données**
- En termes d'**Environnement**, on s'intéresse à un département précis d'un Hôpital et à une maladie spécifique (mélanome) sur une période longue.
- Ceci implique de prendre en compte différentes activités au sein de l'hôpital et différentes réactions des patients.
- Le scénario BI est le scénario 2 : la BI est utilisée pour le suivi de la performance stratégique de l'entreprise.

Santé

- En termes de **Perspective**, on considère les trois possibles. Le responsable du processus est le Département de Dermatologie, les sujets du processus sont les patients et les acteurs sont le personnel de l'hôpital.
- En termes de **Vue**, on utilise les trois vues possibles, selon les objectifs analytiques envisagés. Pour mesurer l'efficacité des check-up et du traitement, on utilise la vue transversale, pour le temps de survie la vue par états et pour l'analyse du processus la vue par événements.

Santé

- En termes d'objectifs analytiques, on se base sur la formulation initiale des KPI pour effectuer :
 1. Une mesure de l'efficacité des check-up préventifs
 2. Un modèle de régression pour la durée de survie des patients
 3. Une analyse de conformité du processus de traitement
 4. Une exploration (mining) de l'organisation du personnel

Santé

- **Activité d'évaluation des données** : chaque objectif analytique requiert un jeu de données spécifique qui est obtenu par intégration et transformation des données initialement disponibles ; par exemple, en effectuant des résumés transversaux de la vue par événement. Une description des données doit être effectuée pour chaque jeu de données et la qualité des données doit être évaluée.
- Dans ce contexte, deux thèmes importants sont l'imputation des valeurs manquantes et l'amélioration de la résolution temporelle (granularité).

Santé

- La deuxième application à la Santé considère un problème spécifique de complication pendant la grossesse : la pré-éclampsie (type d'hypertension).

Santé

- **Entreprise** : Prévention et traitement de la pré-éclampsie de la grossesse. Afin de détecter la pré-éclampsie, le poids, la tension artérielle et la protéinurie des femmes sont surveillées au cours de la grossesse.
- **Objectifs** : On se donne deux objectifs ;
 1. La **représentation** du processus de surveillance par des **mesures quantitatives**
 2. La mise au point de **règles permettant de décider** de l'hospitalisation d'une personne

Santé

- **Activité data** : Des séries temporelles de mesures du poids, des pressions artérielles systoliques et diastoliques et de protéinurie ont été obtenues pour 300 femmes enceintes, dont 8% présentaient une pré-éclampsie.
- **Activité de compréhension de l'entreprise et des données.** En ce qui concerne l'**Environnement**, la portée du processus est définie par six types d'événements : Mesure de protéinurie, mesure de pression artérielle, mesure du poids, hospitalisation, prise en charge à domicile, accouchement. Le scénario BI est le scénario 1 : **BI séparée du management stratégique**, mais les résultats de l'analyse peuvent être utilisés pour mettre au point de nouvelles stratégie de surveillance de la grossesse.

Santé

- En termes de **Perspective**, on retient la **perspective client**. Le responsable du processus est la patiente, qui est également sujet du processus. Les acteurs sont les praticiens médicaux qui décident de l'hospitalisation.
- En termes de **Vue**, on est face à un exemple typique de **vue par états** : on dispose de séries temporelles de paramètres médicaux.

Santé

- On définit comme KPI l'hospitalisation (détection suffisamment tôt du risque de pré-éclampsie) en fonction de l'évolution temporelle des paramètres médicaux. Les objectifs analytiques sont donc **descriptifs** (reporting) et **prédictifs** (règles de classification permettant de prendre une décision d'hospitalisation).
- En termes **d'évaluation des données**, on peut dire que celles-ci sont **exhaustives**, ce qui permet d'évaluer divers algorithmes de classification. Pour ce qui est du reporting, des résumés statistiques standards des instances sont calculés.

Enseignement Supérieur

- Le deuxième domaine d'application de la BI considéré comme exemple est celui de **l'enseignement supérieur**.
- On dispose d'un jeu de données sur des processus affectant différents cours de première et deuxième année de Licence dans une Université.
- Ces cours font intervenir différentes parties prenantes et impliquent un grand nombre de sujets (étudiants) et de participants (enseignants).

Enseignement Supérieur

- **Entreprise** : Etudier le déroulement de cours de première et deuxième année de licence dans une Université à partir de l'usage d'une plateforme d'enseignement en ligne.
- **Objectifs** : On souhaite
 1. Obtenir un modèle de processus de référence, vérifiant des contraintes de conformité basées sur des étapes et des soumissions de devoirs.
 2. Vérifier la conformité entre le processus d'enseignement réel et le modèle de référence.
 3. Mesurer l'utilisation du forum par les étudiants et évaluer la relation entre la performance des étudiants et leurs activités sur le forum.

Enseignement Supérieur

- **Activité data.** Les données ont été collectées à partir de quatre services différents : le forum, les soumissions, les inscriptions et les évaluations, sur la plateforme Web. Des journaux (ang. logs) sur les cours offerts chaque année sur une période de trois ans étaient disponibles. Au total, il y a eu 330 étudiants et 18 511 événements. Les données ont été mises au format .csv et anonymisées.

Enseignement Supérieur

- **Activité de compréhension de l'entreprise et des données.** En termes d'**Environnement**, l'entreprise concerne une faculté.
- La **portée** du processus est complexe en raison des contraintes qui pèsent sur les cours des deux premières années d'université. Un exemple de contrainte est le suivant : « pour chaque étape, aucune soumission ne peut avoir lieu après une date limite fixée ».
- Un **processus de référence** est élaboré à partir d'interviews avec les enseignants.
- Le scénario BI est le scénario 3 : surveiller un processus avec **feedback** sur la stratégie d'entreprise de la faculté. A **long terme**, les résultats peuvent être utilisés comme ressource stratégique pour améliorer les cours (scénario 4).

Enseignement Supérieur

- En termes de **Perspective**, les trois perspectives BI doivent être prises en considération. Le responsable du processus est la faculté et les sujet du processus sont les étudiants. D'autres acteurs sont les enseignants.
- En termes de **Vues**, on a recours à **la vue par événements** pour la compréhension de l'entreprise et à **la vue transversale** pour ce qui est des performances des étudiants.

Enseignement Supérieur

- En termes d'**objectifs analytiques**, on souhaite :
 1. Comprendre le processus dans sa globalité
 2. Effectuer un analyse de conformité des instances du processus
 3. Segmenter les étudiants en différents groupes selon leur utilisation du forum

Enseignement Supérieur

- **Evaluation des données.** Les données sont extraites de l'environnement d'apprentissage et intégrées selon les objectifs d'analyse. Après le **nettoyage des données**, celles-ci sont présentées sous la forme de logs. A l'aide de **transformation des données**, des attributs sont générés pour la vue transversale.

Logistique

- Le troisième domaine d'application de la BI que nous allons envisager est la logistique, plus précisément **le transport de containers** dans le cadre d'une chaîne du froid. Les sujets du processus ne sont donc pas humains (il s'agit des containers).
- Les données utilisées sont des données simulées.

Logistique

- **Entreprise.** On charge un véhicule avec un container et on le suit lors de son déplacement vers sa destination. Durant le déplacement du container, la température est constamment mesurée. Si elle excède un seuil fixé pendant un certain temps, le véhicule doit revenir à son point de départ avec le container. Sinon, il continue vers sa destination où le container est déchargé.

Logistique

- **Objectifs.** On poursuit deux objectifs :
 1. Obtenir un modèle du processus de transport
 2. Obtenir une règle de décision permettant de décider de renvoyer un container vers son point de départ avant livraison.

Logistique

- **Activité data.** On utilise des données simulées de transport et les séries temporelles de mesure de température correspondantes.
- **Activité de compréhension de l'entreprise et des données.** En termes d'**Environnement**, la situation est générale et la portée limitée à un petit nombre d'activités. Le scénario BI est le scénario 1 : La BI est séparée de la stratégie.

Logistique

- En termes de **perspective**, on adopte **la perspective client**. Les containers sont considérés comme **les clients du processus de transport**. Le responsable du processus est la compagnie de logistique et les sujets du processus sont les containers. D'autres acteurs sont les employés de la compagnie.
- En termes de **vues**, on adopte **la vue par événements** pour le premier objectif et **la vue par états** pour le deuxième.

Logistique

- En termes d'**objectifs analytiques**, on souhaite :
 1. Identifier le processus
 2. Classifier les différentes instances du processus de façon à ce qu'une règle de décision de retour économiquement viable puisse être formulée. Le seul facteur influent est la température du container.

Logistique

- **Evaluation des données.** S'agissant de données simulées, celles-ci sont dans un format adéquat et de bonne qualité.

CRM

- Le quatrième et dernier exemple d'application qu'on va envisager est celui de la CRM (Customer Relationship Management, fr. Gestion de la Relation Client).
- Les données utilisées sont réelles.

CRM

- **Entreprise.** Une compagnie avec des magasins dans différentes villes offre une diversité de services à ses clients.
- Les clients sont enregistrés à l'aide d'une carte de fidélité.
- Afin de faire évoluer ses services selon les besoins des clients et d'obtenir de l'information sur l'image de la compagnie, une enquête de satisfaction est effectuée par la compagnie.

CRM

- Environ 2300 clients participent à l'enquête.
- Pour chaque client, l'âge, le sexe, la durée de la relation client et le type d'usage (privé ou professionnel) des services sont obtenus.
- Pour les clients les plus anciens, un profil d'utilisateur est défini à l'aide de trois indicateurs : montant total des ventes, intensité de l'usage, durée de la relation.
- L'utilisation des services est connue à partir de la base de données de transactions.

CRM

- **Objectifs.** Les KPI considérés sont les suivants :
 1. Les ventes aux clients en fonction de leur profil
 2. L'utilisation des différents services
 3. Le profil des clients relativement à leur utilisation des services

CRM

- **Activité data.** On a recours à trois sources de données : une base de données de clients (carte de fidélité), une base de données de transactions et une base de données d'enquête.
- **Activité de compréhension de l'entreprise et des données.** On considère une compagnie de taille moyenne avec un grand nombre d'activités dans différents services, qui sont assez indépendantes les unes des autres. Le scénario BI est le scénario 3 : **La BI rétroagit sur la stratégie de l'entreprise**, voire le scénario 4 : La BI est une ressource stratégique pour la redéfinition de l'offre de services.

CRM

- En termes de **perspective**, on adopte **la perspective client** car l'accent est mis sur les ventes de services. Le responsable du processus est la compagnie et les sujets sont les clients. D'autres acteurs du processus de vente sont les employés de la compagnie.
- En termes de **vues**, on dispose ici de données transversales.

CRM

- Les **objectifs analytiques** découlent des KPI. Il s'agit donc d'abord de segmentation et de classification. D'autres objectifs analytiques sont la description des clients et la détection de comportements atypiques.

CRM

- **Evaluation des données.** Afin d'obtenir des données transversales, on agrège les données transactionnelles à différentes résolutions temporelles (mensuelle, trimestrielle, annuelle).
- La description des variables de vente et l'identification de valeurs aberrantes sont ensuite effectuées, ainsi que la description des résultats de l'enquête de satisfaction client.

Références

- *Fundamentals of Business Intelligence*, W. Grossmann, S. Rinderle-Ma, Springer 2015.
- *Business Intelligence*, R. Sharda et al., Pearson 2013.
- *Key Performance Indicators*, D. Parmenter, Wiley 2015.
- *Designing Performance Measurement Systems*, F. Franceschini et al., Springer 2019.
- *Management by Measurement*, F. Franceschini et al., Springer 2007.